適応型単語リストを用いた自律学習支援システムの構築

Personalized Teaching Material Generator Based on Word Set
- Web ページの語彙レベルの自動測定 –

堀江　郁美*
Ikumi Horie
Email: horie@dokkyo.ac.jp

近来，Web に関する技術の発達により様々な文章が Web ページ化され，容易に閲覧することができるようになった．それらの膨大な Web ページを用いて言語習得や専門知識習得のために学習したいという要望があるが，自分の学習レベルにあった Web ページを見つけるのが困難なため，利用できずにいる学習者も多い．そこで，本研究では，外国語としての英語学習者を対象に，Web ページの語彙レベルを測定するシステムを開発した．これは，英語学習者のライティングに対する評価指標を参考にして，難しいレベルの語彙が使われれば使われるほどテキストを読む難易度が高くなるという考えを基に，語彙リストの General Service List(GSL)と Academic Word List(AWL)を用いて，Web ページの語彙レベルを計算したものである．これによって，グレード別リーディングテキストのように Web ページが学習者にあった語彙レベルであるかどうかを判断できるようになる．本研究では，この語彙レベルを一般の Web ページに適用し実験を行い，Web ページの語彙レベルを計算することによって，学習者の支援ができることを確認した．

Recently, a lot of information is available as web pages on the Internet. Everyone can access and get them easily, and many students want to use them as textbooks. However, it is very difficult for novice learners of English to utilize them as textbooks. They need a help to find the web pages, which are appropriate to their ability. In order to solve this problem, I develop a new system for novice learners of English, which estimates the vocabulary level of the web pages by General Service List(GSL) and Academic Word List(AWL). This system provides suggestions to the students, and they can utilize the web pages as textbooks. Here, I verify that the system is helpful for the students to find the suitable web pages.

*：獨協大学経済学部

1. はじめに

近来，Web に関する技術の発達により様々な文章が Web ページ化され，容易に閲覧することができるようになった．外国語習得や専門知識習得のために，それらの膨大な Web ページを用いて学習したいという要望があるが，自分の語彙レベルにあった Web ページを見つけることが困難なため，利用できずにいる学習者も多い．そこで，本研究では，外国語として英語を学習する学生を対象に，学習者の読みたいと思う Web ページの語彙レベルを測定するシステムを開発した．これによって，学習者は，Web ページを読まなくても，その Web ページが学習者の語彙レベルに適しているかどうかを瞬時に判断できるようになる．

外国語の習得のための学習において，特にリーディング分野では語彙の習得が非常に重要な役割を占めている．英語を外国語とする話者は高頻度語として約 2,000 語のワードファミリー，アカデミック用語として約 600 語のワードファミリーを学習した場合，一般の雑誌，新聞，小説，アカデミック論文などで使用される語彙 80%程度を補うことができることが知られている[1,2]．そこで，英語を外国語とする学習者のために様々な語彙学習方法が研究されている．その中でも，リーディングに関しては，グレード別リーディングテキストの多読を通して語彙学習を行う方法が最も効果的な手法の一つとされている．そこで，本研究では多読のグレード別リーディングテキストの様に，Web ページ中で利用されている語彙レベルによって数値化し学習者に指標として示すことにした．これにより，Web ページを多読のグレード別リーディングテキストの様に扱うことができるようになると期待できる．

語彙学習のための多読においては，語彙のレベルで計算された適切な語彙レベルのテキストを大量に読むことが必要とされている．多読による語彙学習で得られる効果としては，語彙の成長と流暢さの発達が目的とされている．語彙の成長のためにはテキスト内に未知語が 5%未満，流暢さの発達のためには未知語は 0 %がよいとされている．また，未知語を繰り返し学習することは非常に効果的であり，グレード別リーディングテキストによる学習には一定の効果が認められている[1,2]．そこで，本システムでは，学習者に対して含まれる語彙のレベルだけでなく，Web ページの未知語の割合にも着目し，Web ページのカバー率も表示している．これは，先行研究で開発したシステムで採用された機能であり，継続して用いることにした[3,4,5]．

Web ページの語彙レベルの計算には，自由英作文の評価手法を参考にし，難しいレベルの語彙が使われれば使われるほどテキストを読む難易度が高くなるという考えを基に Web ページの語彙レベルの算出を行った．ここでは，語彙のレベル分けのために，標準的に利用されている語彙リストである General Service List (GSL)[6,7] とアカデミック用語リストである Academic Word List(AWL)[8,9]を用いた．GSL の語彙を頻度順に約 200 ワードファミリー単位の 11 レベルに，AWL の語彙を 10 レベルにわけて利用した．これらを用いて，対象となる Web ページの一般的な語彙リストからみた語彙レベルを計算した．これによって，学習者が学習者の語彙レベルにあった Web ページをより適切に選択できるようになった．

先行研究としては，e-learningシステムを採用した語学学習サイトや，語彙をチェックするシステムやゲーム機器を用いた単語学習ゲームソフトなどが多々存在する．語学学習サイトには，英単語を覚えるための単語帳サイトや，Web教材の英単語を抜き出すサイト[10]などがある．しかし，これらのサイトは掲示板や単語ゲームなど多数の機能を持つ非常に優れた語学学習サイトであるにも関わらず，学習者の語彙レベルにあわせたコースを柔軟に選択できない．本研究で開発するシステムでは，利用者が作成する語彙集や定評のある語彙リストを用いて利用者に適したWebページを選択できる．他に，学生に人気の気軽に利用できるサイトに，翻訳サイトがある[11,12]．これらの翻訳サイトは，語学の授業の予習などによく使われているが，本研究が目的としているのは翻訳ではなく，文章読解の学習支援である．L. Anthony[13] は，AWLとGSLに存在する単語がテキスト内にどれだけ含まれているかを示すシステムを作成した．本研究では，AWLとGSLに存在する単語の分布から語彙レベルを算出し利用者に提示できるため，Webページ同士の比較も可能である．S. Haywood[14]は，テキスト中でAWLに存在する単語をハイライトするシステムを開発した．非常に有益なシステムではあるが，本研究では，単語リストを作成したり，選択できたり，Webページの語彙レベルを提示したりとより柔軟に処理できる．

本論文の構成は以下のとおりである．2 章で外国語としての英語学習について紹介する．2.1 章では英語学習の中での語彙の役割，2.2 章でリーディングと語彙との関係，2.3 章で語彙リストについて，2.4 章では Web ページの語彙レベルを計算するのに参考にした自由英作文評価手法について述べる．3 章では，提案するシステムについて詳しく説明する．3.1 章で Web ページの語彙レベルの計算の仕方，3.2 章では評価実験について延べ，3.3 章では提案するシステムについてまとめる．4 章では本研究をまとめる．

2. 外国語としての英語学習

外国語習得のための学習方法の中で，グレード別リーディングテキストの多読を通して語彙学習を行う方法は重要な手法の一つである．ここでは，語彙とリーディングの関係について詳しく説明した後，自由英作文を用いた評価指標について詳しく述べる．

2.1 語彙について

語彙知識とリーディング力は相互に非常に密接に関連することがわかっている．ここでは，英語を母国語

とする話者や，外国語とする話者がどの程度の語彙数を必要とするかなどを用いて，語彙習得が難易度を表す一つの指標と得ることを説明する．

まず，以下に３つの単語の数え方を説明する．

(1) **語種数(Type)**：使用される単語の個数．但し，同じ単語が２回以上生じる場合，繰り返し数えない．
(2) **延べ語数(Tokens)**：使用される単語の個数．同じ単語が２回以上生じた場合も，繰り返し数える．
(3) **ワードファミリー(Word family)**：見出し語とその屈折系および密接に関連する派生語からなる．

英語を母国語とする話者は約 20,000 ワードファミリーを知っていると示唆されている[1]．これに対して，英語を外国語とする話者は，約 3,000 程度のワードファミリーが必要とされる．表 1．では，会話，小説，新聞，アカデミックの分野で高頻度語や，アカデミック用語におけるテキストのカバー率を示している．

Levels	Conversation	Fiction	Newspapers	Academic
1st 1000	84.3	82.3	75.6	73.5
2nd 1000	6	5.1	4.7	4.6
Academic	1.9	1.7	3.9	8.5
Other	7.8	10.9	15.7	13.3

表 1．様々な種類のテキストのカバー率[1,2]

語彙は，次の様に，高頻度語，低頻度語，専門用語，アカデミック用語の４つにわけられる[1]．

(1) **高頻度語**：前置詞や接続詞などの機能語も含め，高頻度で出現する単語．テキストに含まれる総語数のうち約 80%は高頻度語である．
(2) **アカデミック用語**：アカデミックな教科書から抜き出されたもので，異なる科目のテキストにも共通する単語．テキストの総語数の約 9%をしめる．
(3) **専門用語**：その主題および対象分野と非常に密接な関係がある単語．対象分野では高頻度で現れるが，それ意外の分野ではあまり出現しない．通常はテキストの総語数の 5%程度であるが，対象分野によって異なる．
(4) **低頻度語**：高頻度語，アカデミック用語，専門用語に含まれない単語．アカデミックなテキストの単語の 5%以上となる．

表２は，総語数が 500 万語のテキスト集における語彙のカバー率を示している．Nation は，約 2,000 語の高頻度語で 80%以上のテキストをカバーしており，次がアカデミック用語，専門用語，低頻度語の順でカバーしていくことを示した[1]．そこで，高頻度語，アカデミック用語，専門用語，低頻度語の順で学習することが推奨されている．これにより，テキストなどのリーディングの語彙による難易度も，上記と同じ順序で並ぶことが想像できる．高頻度語の語彙の学習方法には，Direct teaching, Direct learning, Incidental learning, Planned encounters が推奨されており，多読リーディングによる語彙学習は，Planned encounters に分類される．

Number of words	% text coverage
86,741	100
43,831	99
12,448	95
5,000	89.4
4,000	87.6
3,000	85.2
2,000	81.3
1,000	74.1
100	49
10	23.7

表 2. Vocabulary size and coverage [1]

2.2　リーディングと語彙

外国語習得の中で，語彙の学習方法として多読が有効であることがわかっている[1,2]．多読では，テキストが学習者の語彙レベルに合致している必要があるためグレード別リーディングテキストというものがよく利用される．また，リーディングの目的によってテキスト内の未知数の割合が95～99%であることが必要とされている．これらについて順に説明する．

リーディングは，テキストを綿密にじっくりと学習する精読と，適切な語彙レベルで大量の読書を行う多読に分類される．精読においては，約 300～500 語程度の短いテキストを語彙，文法，内容に関して，テキストに付随する注釈や，語彙エクササイズ，テストを用いて学習することが効果的であると言われている．多読においては，同じ語彙に繰り返し出会う機会を増やすことが必要とされ，学習者が理解可能なテキストを大量に読むことによって語彙の習得を可能とする．精読にも多読にも，学習者の語彙レベルにテキストがあっていることが重要視されている．そこで，学習者の語彙レベルとテキストの語彙レベルをあわせるために，多くの場合，グレード別リーディングテキストが用いられる．グレード別リーディングテキストを用いた多読は一定の効果があることが示されている．

グレード別リーディングテキストとは，厳密に限定された語彙レベルの範囲内で作成された読みものであり，通常いくつかの語彙レベルに分割される．表3．はグレード別リーディングテキストの一つであるペンギンリーダーズの語彙レベルと TOEIC，TOEFL，英検の点数の関係である．ペンギンリーダーズでは，７つの語彙レベルに分かれており，Easystarts レベルのテキストは高頻度語の 200 語の語彙だけで読めるようになっている．この語彙レベルのテキストは，200 語の語彙レベルにするために非常に簡易化された文章となっており約 20 ページほどである．読者は次のレベル１では 100 語追加し 300 語の語彙で読むことができる．このように，段階を追って順に語彙レベルを増やすことが

できる上に，語彙がコントロールされているので，学習者の語彙レベルに合わせることが可能である．

レベル	見出し語数	TOEIC	TOEFL iBT	英検
Easystarts	200	250	26-27	4級
レベル1	300	250	26-27	4級
レベル 2	600	350	36-37	3級
レベル 3	1,200	400	40	準2級
レベル 4	1,700	500	52	2級
レベル 5	2,300	600	62-63	2級-準1級
レベル 6	3,000	730	79-80	準1級

表3．グレード別リーディングテキスト(ペンギンリーダーズ)のレベルとTOEIC，TOEFL，英検の関係 [15]

また，英語の言語習熟度についての研究では，興味のある内容のリーディングが習熟度と最も強く相関があり，授業外のリーディングがTOEFLテストの結果に最も重要で直接的に寄与することがわかっている[1]．これらが，文法などの語学学習用テキストの他に，様々なテーマを扱ったグレード別リーディングテキストが効果的であるとされている理由である．

次に，多読用テキストと未知語のカバー率について説明する．多読の場合は，流暢さを発達させ語彙知識の深さを増加させるのか，語彙知識の幅を増加させるのか目的にあわせて，テキスト内に含まれる単語のカバー率を変える必要がある[1，2]．例えば，カバー率が98%以上の場合，学習者にとって読みやすいテキストとなりリーディングを楽しむことができるが，ほとんど未知語がないため語彙量を増加させることができず，リーディングの流暢さを促すのみとなる．次に，カバー率が95%から順に下がるとだんだん適切な理解ができる学習者は減少し，80%で適切な理解のできる学習者はいなくなる．このため，少なくともカバー率は80%以上である必要があり，流暢さの発達のためにはカバー率98～99%，語彙の増加を目的とするには95%のカバー率が推奨されている．

2.3 語彙リストについて

ここでは，カバー率を測るために利用する語彙リストについて説明する．

現在様々な語彙リストが存在し，大きくわけて標準語彙リスト，アカデミック語彙リストが存在する．代表的な標準語彙リストには表 4 で示されるように，GSL(General Service List), Teacher’s Word Book of 30,000 Words , Computational Analysis of Present-Day American English, Cambridge English Lexicon, Word Frequencies in Written and Spoken English が存在する．これらは，テーマを問わず様々な分野の雑誌や書籍の中で高頻度に出現する単語をリスト化したものである．これらの語彙リストの中で，使用頻度，ある単語リストがテキストに締める占有率，どれほど多くの異なるテキストに出現するかという使用範囲，覚えやすさや単語の関連度などを考慮にいれた教育的配慮から判断されるのは GSL である．また，欧米の出版社からだされたグレード別リーディングテキストは GSL を用いて書かれたものが多く，現在，多くの語彙学習のテキストで利用されている．

アカデミック語彙リストには，表 5 のように UWL(University Word List)，AWL(Academic Word List) ，English Vocabulary for Academic Purposes などが存在する．これらは GSL のような基本語彙と異なり，一般的なテキストではあまりみかけないが，アカデミックなテキストにおいて頻繁に出現する単語のリストである．この中でも AWL が最良のリストと呼ばれ，GSL の約 2000 語の基本語彙を既に知っているものと仮定して，次に習得すべきリストとしてよく利用されている．

この様に，現在では数ある語彙リストの中で，語彙学習の順序としては，使用頻度の高い単語から学習することが効率的であるため，GSL を高頻度順に 2000 語を学習した後に，AWL を学習することが理想とされている．そこで，本研究では，GSL と AWL を利用して語彙レベルを算出することにした．

名称	語数	発行年	備考
GSL(General service List)	高頻度 2000 語	1953	英語学習用，word-family 方式
Teacher’s Word Book of 30,000 Words	頻度別 30,000 語	1944	見出し語方式
Computational Analysis of Present-Day American English	5,000 語	1967	見出し語方式
Cambridge English Lexicon	約 4,500 語	1980	指導者用，見出し語方式
Word Frequencies in Written and Spoken English	100,000,000 語	2001	Spoken と Written に区別，品詞別による頻度

表4．代表的な標準語彙リスト [2]

名称	語数	発行年	備考
UWL(University Word List)	基本語 836 語と派生形 1,400 語	1984	word-family 方式
AWL(Academic Word List)	570 語と 10 の補助リスト	2000	word-family 方式
English Vocabulary for Academic Purposes	7,692 語	1997	見出し語方式

表5. アカデミック語彙リスト[2]

2.4 ライティング評価手法

本研究では，文章の難易度は文章中の語彙の難易度に比例するという考えを基に文章の難易度を測定している．同様の考えを用いた評価手法として，外国語としての英語学習者のライティングに対する評価指標の一つに，語彙(vocabulary)が挙げられる．Web ページの語彙レベルを測定するために，ライティングの評価指標を参考にした．

米国の GMAT(Graduation Management Admission Test)の小論文の採点で使用されている e-rater は人間の評価との一致度が97%であると報告されている[16]. この他にもライティングを自動的に評価しようという研究は多くされている．これらでは主に以下の項目を用いて評価基準を作成している[16,17,18].

(1) Type-token ratio (TTR)：語種数と延べ語数の比率を表す
(2) Standardized Type/Token Ratio(標準化 TTR：TTR を標準化したもの
(3) Guiraud：語種数を延べ語数の平方根で割った値
(4) Mean Word Length：平均語長
(5) Setences：センテンス数
(6) Mean Words/Sentences：センテンスあたりの平均語数
(7) Lexical density(LD)：内容語と機能語の比率を表す
(8) Lexical Frequency Profile(LFP)：使用語彙のレベルを測定

この様に，ライティングの自動採点システムにおける説明変数のうち語彙の締める役割は大きい．上記 8 個の項目は自由英作文の指標ということで，本来は他に文法や熟語など他の要因も関係するが，本研究では既にできあがったテキストであることを考慮し，Web ページ内に文法間違いなどはないとみなし語彙のみに注目した．

TOEIC 模擬テストの習熟度を測るのにライティングの自動採点手法を用いた研究では，これらのうち延べ語数と LFP，標準化 TTR が順に相関が高かった．LFP の指標の一つに，次の式の様に，難しいレベルほど難易度の高い語彙を使用しているという考えを基にレベルで重み付けを行い計算を行うものがある．

LFP = $\Sigma_{n=1}^{k}$ レベル n * テキスト内の n レベルの語の語数

これは，語彙リストをいくつかのレベルに分類し，そのレベルを表す数字と，そのレベルで使用されている語彙の数をかけたものを全てのレベルで計算し足したものである．延べ語数が多ければ多いほど数値が高くなることに注意する必要がある．

本研究では，この指標を参考にし，Web ページの語彙レベルを計算し，学習者にあった語彙レベルの Web ページを見つけることができるようにした．次に本研究で用いた指標の計算方法や実験結果を示す．

3. システムと評価実験

以上のことを参考にし，システムを作成し，評価実験を行った．3.1 章では，用いた語彙リストや指標について，3.2 章では評価実験について，3.3 章では評価実験をまとめる．

3.1 評価システム

ここでは，使用する語彙リストと指標について説明する．

使用する語彙リストについて

語彙リストとしては，先行研究を参考にして GSL と AWL を用いた．これは，歴史的に評価が高いことや，表 7 の様に多くの一般的グレード別リーディングテキストと同様に様々な分野の単語を父君でおり，カバー率が高いことが理由である．

GSL には 2284 のワードファミリーが含まれるが，頻度順に約 200 ワードファミリーずつ 11 のレベルに分割[1]した後，ワードファミリーをそれぞれの単語に戻した. 例えば, ワードファミリー中の単語 research は，最終的には researched，researcher，researchers，researches， researching の 5 単語に分解される．

最終的に，単語数は全部で 7,540 語あり，1 つのレベルには平均して約 650 の語彙が含まれた(表 8). AWL では, 570 のワードファミリーが既に 10 のレベルにわかれているものを用いた．表 9 の様に，ワードファミリーを分解することによって 3,113 語になった．本研究では，GSL を学習した後に AWL を学習するとみなして，語彙レベルを GSL の高頻度順から AWL の順に順序付けを行った．

Lists	Fiction	Popular	Newspapers	Academic
Readers	85.6%	81.9%	80.1%	76.3%
GSL	85.5%	81.5%	79.7%	76.4%
GSL+AWL	86.4%	86.4%	84.9%	85.1%

表 7. 一般的なグレード別リーディングテキストであるリーダーズと GSL, AWL の関係[1]

[1] 最後のレベルのみ 284 ワードファミリーとなった.

レベル	General Service List, level	単語数
1	General Service List, level 1	763
2	General Service List, level 2	840
3	General Service List, level 3	876
4	General Service List, level 4	802
5	General Service List, level 5	747
6	General Service List, level 6	687
7	General Service List, level 7	592
8	General Service List, level 8	592
9	General Service List, level 9	618
10	General Service List, level 10	529
11	General Service List, level 11	494
	合計	7,540

表 8. General Service List のレベルと含まれる単語数

レベル	Academic Word List, level	単語数
12	Academic Word List, level 1	453
13	Academic Word List, level 2	395
14	Academic Word List, level 3	366
15	Academic Word List, level 4	293
16	Academic Word List, level 5	339
17	Academic Word List, level 6	336
18	Academic Word List, level 7	267
19	Academic Word List, level 8	309
20	Academic Word List, level 9	260
21	Academic Word List, level 10	95
	合計	3,113

表 9. Academic Word List の語彙レベルと含まれる単語数

テキストの語彙レベルを表すための指標について

本研究では，Web ページの語彙レベルを計算するために，難しいレベルの語彙が使われれば使われるほどリーディングの難易度が高くなるという考えを基に重み付けを行い次の指標を作成した．これを，一般的なテキストの標準的語彙レベル(Standard Vocabulary Level)と呼ぶ．

語彙リストとして表 8, 表 9 の様に分解された GSL と AWL を用いた．GSL を頻度順に 1 から 11 レベル，そして，AWL を GSL を学習した後に習得すべきリストと解釈し，12 から 21 レベルとした．以下で，n は語彙レベルを表す数字とする．

テキストの標準的語彙レベル(SVL)

SVL はリーディングテキストがどの語彙レベルの語彙で作成されているかを示す指標で，次の様な式で表される．

$$SVL = \Sigma_{n=1}^{21} \text{語彙レベル} n \ * \ \text{Web ページ内の} n \text{ レベルの語の語数} \div \text{述べ語数}$$

LFP では，使用される延べ語数が多ければ多いほど SVL 値は高くなった．よって，同じレベルの語彙だけで作成されているテキストであっても述べ語数が多いテキストほど何度は高くなってしまうため延べ語数で割ることでこれを解決した．この指標を用い，Web ページの語彙レベルを測定した．

3. 2 評価実験

以下の 4 つの Web ページを対象として, それぞれの語彙レベルを計算した．システムは，以前作成した Web ページからカバー率を計算し, 未習得単語を自動的に辞書をひくシステムを改良した[7,8,9].

1) 獨協大学 英語版トップページ
2) VOA NEWS
3) The Asahi Shimbun English Web Edition
4) BBC Learning English

1) 獨協大学 英語トップページ[19] : **Spirit of the School Establishment**[19]
（参考資料 1）

語種数 59，リストに存在する単語数 57
カバー率 : 96.61 %
Web ページの語彙レベル: 3.68

レベル	単語リスト	出現数	レベル	単語リスト	出現数
1	GSL1	35	12	AWL1	1
2	GSL2	3	13	AWL2	3
3	GSL3	6	14	AWL3	0
4	GSL4	1	15	AWL4	0
5	GSL5	2	16	AWL5	2
6	GSL6	0	17	AWL6	1
7	GSL7	1	18	AWL7	0
8	GSL8	0	19	AWL8	0
9	GSL9	0	20	AWL9	1
10	GSL10	1	21	AWL10	0
11	GSL11	0			

表 10. 獨協の英語版トップページ内に出現する語彙レベル

2) VOA News : News / Science & Technology NASA Closes In on the Big Bang[20]

語種数 121，リストに存在する単語数 87
カバー率: 71.90 %
Web ページの語彙レベル: 3.01

レベル	単語リスト	出現数	レベル	単語リスト	出現数
1	GSL1	47	12	AWL1	1
2	GSL2	9	13	AWL2	1
3	GSL3	13	14	AWL3	0
4	GSL4	3	15	AWL4	0
5	GSL5	4	16	AWL5	1
6	GSL6	2	17	AWL6	1
7	GSL7	1	18	AWL7	2
8	GSL8	1	19	AWL8	0
9	GSL9	1	20	AWL9	0
10	GSL10	0	21	AWL10	0
11	GSL11	0			

表 11. VOA のある記事内に出現する語彙レベル

3) Asahi Shimbun English Web Edition: SUMO/ First new yokozuna in 5 years inspired by family, mentor[21]

語種数 246，リストに存在する単語数 177
カバー率：71.95 %
Web ページの語彙レベル: 3.05

レベル	単語リスト	出現数	レベル	単語リスト	出現数
1	GSL1	85	12	AWL1	2
2	GSL2	28	13	AWL2	3
3	GSL3	11	14	AWL3	0
4	GSL4	12	15	AWL4	2
5	GSL5	11	16	AWL5	0
6	GSL6	9	17	AWL6	0
7	GSL7	4	18	AWL7	0
8	GSL8	5	19	AWL8	0
9	GSL9	3	20	AWL9	0
10	GSL10	2	21	AWL10	0
11	GSL11	0			

表 12. Asahi Shimbun のある記事内に出現する語彙レベル

4) BBC Learning English: Talk about English The Reading Group Part 10[22]

語種数 599，リストに存在する単語数 473
カバー率: 78.96 %
Web ページの語彙レベル: 4.10

レベル	単語リスト	出現数	レベル	単語リスト	出現数
1	GSL1	173	12	AWL1	7
2	GSL2	88	13	AWL2	10
3	GSL3	45	14	AWL3	6
4	GSL4	35	15	AWL4	4
5	GSL5	33	16	AWL5	5
6	GSL6	20	17	AWL6	3
7	GSL7	11	18	AWL7	6
8	GSL8	7	19	AWL8	7
9	GSL9	6	20	AWL9	2
10	GSL10	5	21	AWL10	1
11	GSL11	1			

表 13. BBC Learning English の記事内に出現する語彙レベル

3. 3 評価実験のまとめ

ここでは，評価実験をまとめる．4 つの Web ページでは，それぞれカバー率と Web ページの語彙レベルが計算された．

	カバー率	語彙レベル
獨協	96.61%	3.01
VOA	71.90%	3.49
ASAHI	71.95%	3.05
BBC	78.96%	4.10

表 14. 各記事内のカバー率と語彙レベル

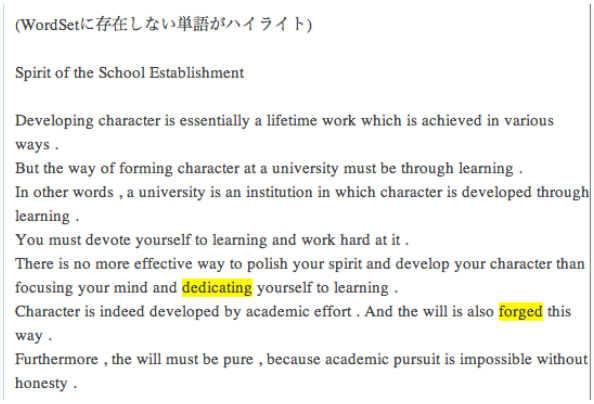
語種数では獨協の 59 語から BBC learning の 599 語まで開きがあったが，延べ単語数に関係なく比較可能なカバー率，語彙レベルの値がでた．

表 14 より，カバー率が一番高いのは獨協の英語版トップページであり，低いのが VOA の記事であった．レベルでは，BBC の記事が一番語彙レベルが高く，獨協の記事が一番低かった．獨協大学の英語トップページのカバー率が高いのは，教育機関のため誰にでも理解しやすい語を選んでいることと，短い文章のため，GSL や AWL に存在する単語で文章が書かれているものと思われる．VOA，ASAHI，BBC の記事でカバー率が比較的低いのは，VOA は天文に関する記事，ASAHI は相撲に関する記事だったので比較的専門用語が多かったからだと思われる．BBC に関しては，テーマがアフリカであったため，Ethiopia や Sierra Leone といった国名や UNICEF といった団体名が未知語としてみなされたことも理由の一因であると思われる．語彙に関しては，BBC で使われている単語は他に比べて比較的 AWL に含まれる単語が多く，英語学習者用のサイトではあるが，少々語彙レベルが高いことがわかる．

獨協大学の英語版トップページでは，カバー率も高く，語彙レベルも低いので，比較的読みやすい Web ページであることがわかる．また，ASAHI の語彙レベルはそんなに高くはないが，カバー率が低いため読む事

が多少困難なことがわかる．リーディングテキストや，言語習得用グレード別リーディングテキストでは，注釈や重要単語リストなどが効果的であるということがわかっている[1, 2]ため，辞書や注釈も学習者に提示し，カバー率をあげるかわりとなる更なる学習支援が必要なページであることもわかった．

これらのことより，自律支援を目的とするシステムとして，Web ページの語彙レベルとカバー率を提示することは外国語として英語を学習する学習者にとって効果があると思われる．但し，英語学習者用のサイトでもカバー率が低いことも多く，カバー率が 95%以下の場合には注釈や重要単語リストなどを提示する更なる支援が必要なことがわかった．実際には，図 1 のように，本研究で使用しているシステムでは，既に，注釈や未習得の単語に関して自動的に辞書をひくなどの支援は行っている．今後は，それら以外に効果的な支援を調べ追加する予定である．


(WordSetに存在しない単語がハイライト)

Spirit of the School Establishment

Developing character is essentially a lifetime work which is achieved in various ways .
But the way of forming character at a university must be through learning .
In other words , a university is an institution in which character is developed through learning .
You must devote yourself to learning and work hard at it .
There is no more effective way to polish your spirit and develop your character than focusing your mind and dedicating yourself to learning .
Character is indeed developed by academic effort . And the will is also forged this way .
Furthermore , the will must be pure , because academic pursuit is impossible without honesty .

図 1. 獨協大学英語版トップページのテキストをシステムにいれ，語彙リストに存在しない単語をハイライトしたもの

また，今回は Web ページの語彙レベルを調べたが，同じリーディングテキストでも，学習者が変われば，カバー率やレベルが変わるはずである．この学習者に対するテキストのレベルは，学習者の習得した語彙をシステムに記録させることによって，容易に計算できる．また，学習者の習得した語彙のレベルも同じようにして容易に計算できるはずである．今後は，学習者の語彙レベルや学習者に対するテキストのレベルを計算し提示するよう改良したい．

また，学習者によって利用したいリーディングテキストなども変わるため，レベルを様々なテキストのレベルに変換できたり，TOEIC や TOEFL の点数に変換できたりと改良する必要がある．

4. おわりに

本研究では，GSL と AWL の語彙リストを用いて，グレード別リーディングテキストのように Web ページを扱うために，Web ページの語彙レベルを測定するシステムを開発した．提案した語彙のレベルと語彙リストへのカバー率を提示することで，自分のレベルにあっているかどうかを判断することができるようになった．

謝辞

本研究の一部は，情報科学研究所研究助成，獨協大学研究奨励費によるものである．

参考文献

(1) I.S.P. ネーション，吉田晴世，三根浩，”英語教師のためのボキャブラリーラーニング”，松柏社，2005
(2) 門田修平編著，池村大一郎，中西義子ほか，“英語のメンタルレキシコン”，松柏社，2004
(3) 堀江郁美，飯島優雅，”学生の自律学習を支援する適応型単語リスト作成ツールの開発”，獨協大学情報科学研究，第 27 号，p59-66，2010
(4) Ikumi HORIE, Kenji KASHIWABARA, Kazunori YAMAGUCHI, Yuka IIJIMA, "Personalized Teaching Material Generator Based on Word Set," Information Technology Based Higher Education and Training (ITHET), 2010, pp. 343-348.
(5) “A Word List Generator Program for Using Authentic Texts in an Academic English Reading Class”, Iijima, Yuka, Horie, Ikumi,, Information Technology Based Higher Education and Training, p. 407-412, 2010
(6) J. Bauman, and B. Culligan, About the General Service List, http://jbauman.com/gsl.html, 1995
(7) A General Service List of English Words, West, M., Longman, 1953
(8) A. Coxhead, A new academic word list, TESOL Quarterly, 34, pp.213-238, 2000
(9) The Academic Word List, http://www.victoria.ac.nz/lals/resources/academicwordlist/
(10) ライフサイエンス辞書プロジェクト, http://lsd.pharm.kyoto-u.ac.jp/ja/index.html
(11) Excite 翻訳，http://www.excite.co.jp/world/
(12) Yahoo 翻訳，http://honyaku.yahoo.co.jp/
(13) L. Anthony, From Language Analysis to Language Simplification with AntConc and AntWordProfiler (Summary of JAECS 2008 workshop), JACET Newsletter, Issue: 63, p.2
(14) Sandra Haywood, AWL Highlighter, http://www.nottingham.ac.uk/%7Ealzsh3/acvocab/awlhighlighter.htm
(15) PEARSON NEW PENGUIN READERS, http://www.longmanjapan.com/penguin/readers_j.html
(16) 杉森直樹，”Lexical Frequency Profile を用いた L2 ライティングにおける語彙的豊かさの評価”，立命館言語文化研究，21 巻 2 号，2009
(17) 杉浦正利，”英文ライティング能力の評価に寄与する言語的特徴について”，学習者コーパスに基づく英語ライティング能力の評価法に関する研究，平成 17 年度-平成 19 年度科学研究費補助金研究基盤 C 研究成果報告書，p33-58，2008
(18) 水本篤，”自由英作文における語彙の統計指標と評定者の総合的評価の関係”，統計数理研究所共同研究リポート 215，学習者コーパスの解析に基づく客観的作文評価指標の検討，p.15-28，2008
(19) “Sprit of the School Establishment”, DOKKYO UNIVERSITY, http://www.dokkyo.ac.jp/english/index_e.html
(20) “NASA Closes In on the Big Bang”, Voice Of America,

http://www.voanews.com/content/hubble-telescope-shows-earliest-galaxies/1515236.html
(21) “SUMO/First new yokozuna in 5 years inspired by family, mentor”, The Asahi Shimbun, http://ajw.asahi.com/article/sports/sumo/AJ201209280012
(22) “BBC Learning English Talk about English The Reading Group Part 10, BBC, http://downloads.bbc.co.uk/worldservice/learningenglish/webcast/readinggroup_prog10.pdf

参考資料 1: 獨協大学英語版トップページ テキスト

Spirit of the School Establishment

Developing character is essentially a lifetime work which is achieved in various ways. But the way of forming character at a university must be through learning. In other words, a university is an institution in which character is developed through learning. You must devote yourself to learning and work hard at it. There is no more effective way to polish your spirit and develop your character than focusing your mind and dedicating yourself to learning. Character is indeed developed by academic effort. And the will is also forged this way. Furthermore, the will must be pure, because academic pursuit is impossible without honesty.

Founder of Dokkyo University-Dr. Amano Teiyu

(2012 年 9 月 21 日受付)
(2012 年 12 月 19 日採録)